

How can I use dplyr to replace all NA values with the mean in a dataset?

Authored by
stats writer

June 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I use dplyr to replace all NA values with the mean in a dataset?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=155186>

Dplyr is a powerful R package that allows for efficient data manipulation and analysis. One useful function of dplyr is the ability to replace missing values, represented as NA, with the mean of the dataset. This can be done using the "mutate" function and specifying the column containing the NA values. The "na.rm = TRUE" argument is also necessary to ensure that the mean is calculated without including the NA values. This feature of dplyr simplifies the process of handling missing data and helps to ensure accurate analysis of the dataset.

Replace NA with Mean in dplyr

You can use the following methods to replace NA values with the mean using functions from the dplyr and tidyr packages in R:

Method 1: Replace NA values with Mean in One Column

```
df %>% mutate(across(col1, ~replace_na(., mean(., na.rm=TRUE))))
```

Method 2: Replace NA values with Mean in Several Columns

```
df %>% mutate(across(c(col1, col2), ~replace_na(., mean(., na.rm=TRUE))))
```

Method 3: Replace NA values with Mean in All Numeric Columns

```
df %>% mutate(across(where(is.numeric),
```

```
~replace_na(., mean(., na.rm=TRUE)))
```

The following examples show how to use each method in practice with the following data frame:

```
#create data frame
```

```
df <- data.frame(player=c('A', 'B', 'C', 'D', 'E'),  
points=c(17, 13, NA, 9, 25),  
rebounds=c(3, 4, NA, NA, 8),  
blocks=c(1, 1, 2, 4, NA))
```

```
#view data frame
```

```
df
```

```
player points rebounds blocks
```

```
1 A 17 3 1
```

```
2 B 13 4 1
```

```
3 C NA NA 2
```

```
4 D 9 NA 4
```

```
5 E 25 8 NA
```

Example 1: Replace NA Values with Mean in One Column

The following code shows how to replace the NA values in the points column with the mean value of the points

column:

```
library(dplyr)
```

```
library(tidyr)
```

```
#replace NA values in points column with mean of  
points column
```

```
df <- df %>% mutate(across(points, ~replace_na(.,  
mean(., na.rm=TRUE))))
```

```
#view updated data frame
```

```
df
```

```
player points rebounds blocks
```

```
1 A 17 3 1
```

```
2 B 13 4 1
```

```
3 C 16 NA 2
```

```
4 D 9 NA 4
```

```
5 E 25 8 NA
```

The mean value in the points column was 16, so the one NA value in the points column was replaced with 16.

All other columns remained unchanged.

Example 2: Replace NA Values with Mean in Several Columns

The following code shows how to replace the NA values in the points and blocks columns with their respective column means:

```
library(dplyr)
```

```
library(tidyr)
```

```
#replace NA values in points and blocks columns with  
their respective means
```

```
df <- df %>% mutate(across(c(points, blocks),  
~replace_na(., mean(., na.rm=TRUE))))
```

```
#view updated data frame
```

```
df
```

```
player points rebounds blocks
```

```
1 A 17 3 1
```

```
2 B 13 4 1
```

```
3 C 16 NA 2
```

```
4 D 9 NA 4
```

```
5 E 25 8 2
```

Example 3: Replace NA Values with Mean in All Numeric Columns

The following code shows how to replace the NA values in every numeric columns with their respective mean

value:

```
library(dplyr)
```

```
library(tidyr)
```

```
#replace NA values in all numeric columns with their  
respective means
```

```
df <- df %>% mutate(across(where(is.numeric),  
~replace_na(., mean(., na.rm=TRUE))))
```

```
#view updated data frame
```

```
df
```

```
player points rebounds blocks
```

```
1 A 17 3 1
```

```
2 B 13 4 1
```

```
3 C 16 5 2
```

```
4 D 9 5 4
```

```
5 E 25 8 2
```

Notice that the NA values in all numeric columns have been replaced with their respective column means.

The one column that was not numeric (player) has remained unchanged.

The following tutorials explain how to perform other common tasks in dplyr:

ARABPSYCHOLOGY.COM